En route to the Clinic: Diagnostic Sequencing Applications Using the Ion Torrent

Donna Muzny M.Sc.
Baylor College of Medicine
Human Genome Sequencing Center

BCM
Baylor College of Medicine

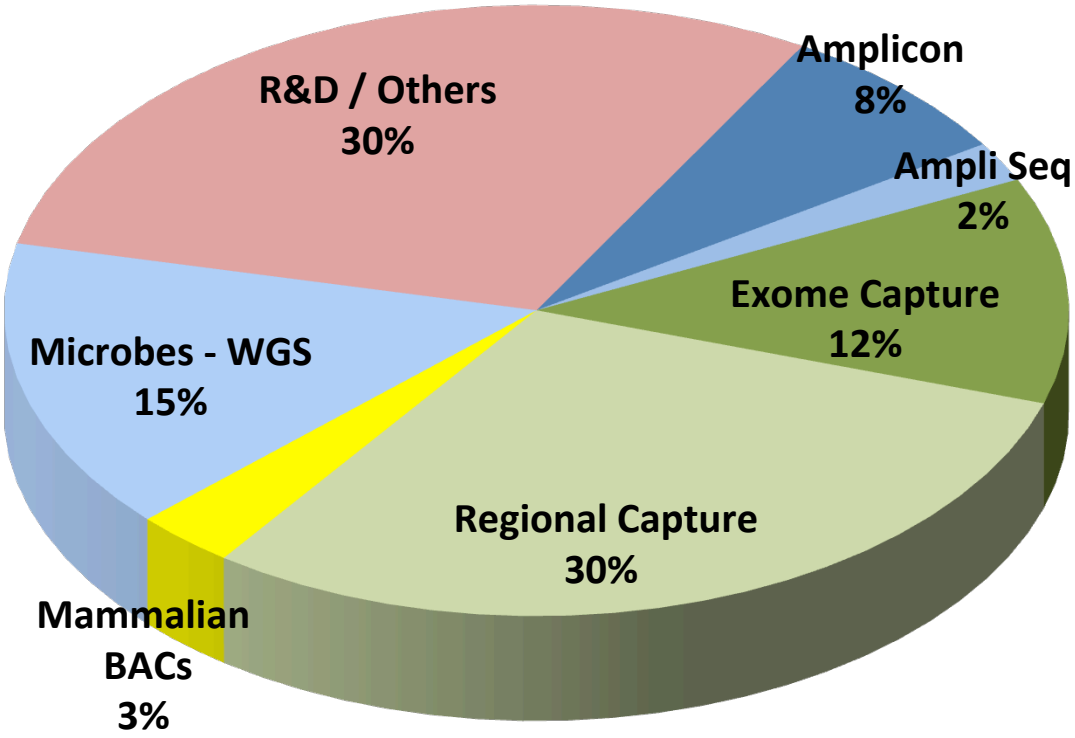
HGSC
HUMAN GENOME SEQUENCING CENTER

Ion Torrent Platform Development

Ion Torrent PGM Activities – Instrument received Jan 2011

- **Microbial Genomes and Mammalian BACs**
  - Platform validation
  - Establish read error rates
  - Assembly consensus error rates and assembly metrics
- **Variant discovery and validation**
  - Amplicon Sequencing
  - Capture Sequencing
- **Technology Advancement**
  - Library construction techniques
  - iShear protocol
  - AmpliSeq Panels
  - 2X bead deposition protocol
  - Improved breaking protocol
  - One Touch system
  - Software releases – Version 2.2 - Improved analysis algorithms for signal processing, basecalling and alignment, Improved error rates

Distribution of Ion Torrent Runs



Total Runs	**426**
Total Gb	**133**

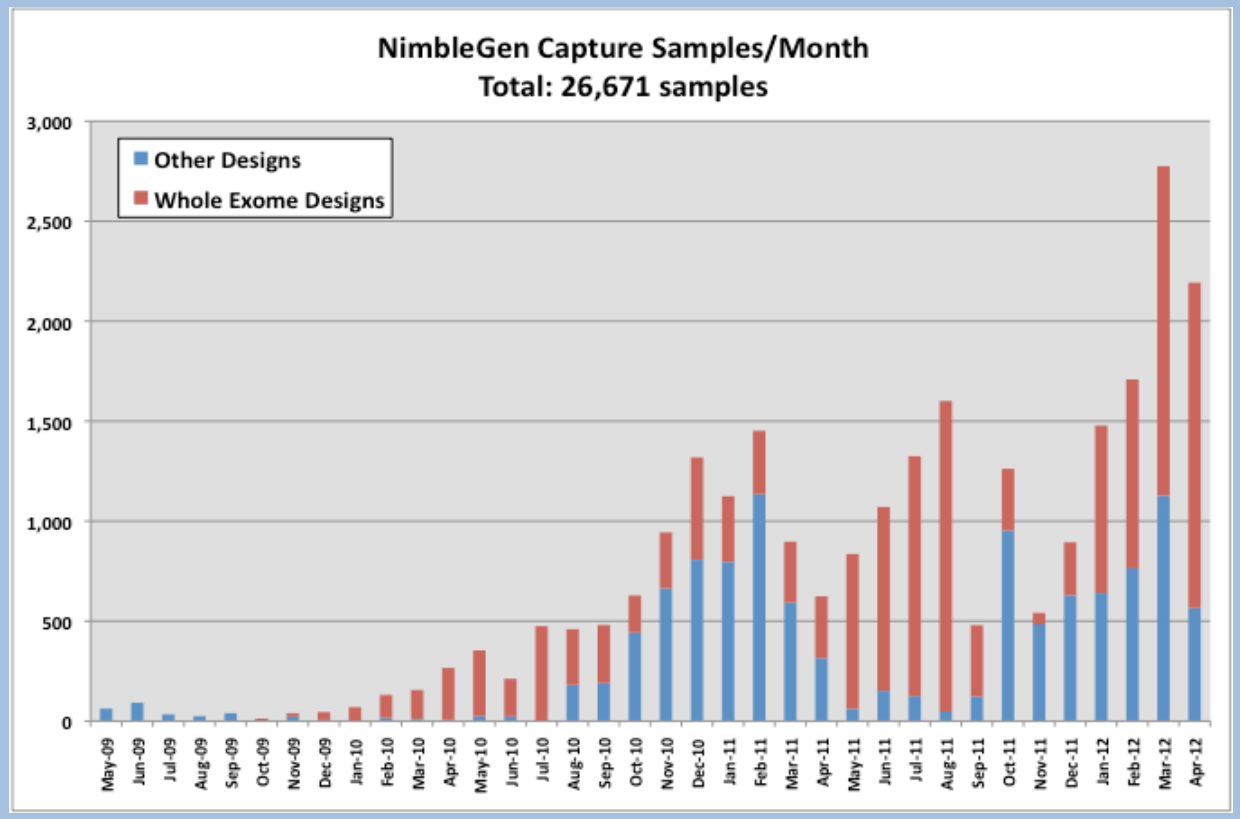
Best In House 316 Run

Pass Filter [Mbp]	613
Aligned Q17 [Mbp]	507
Aligned Q20 [Mbp]	439

Best In House 318 Run

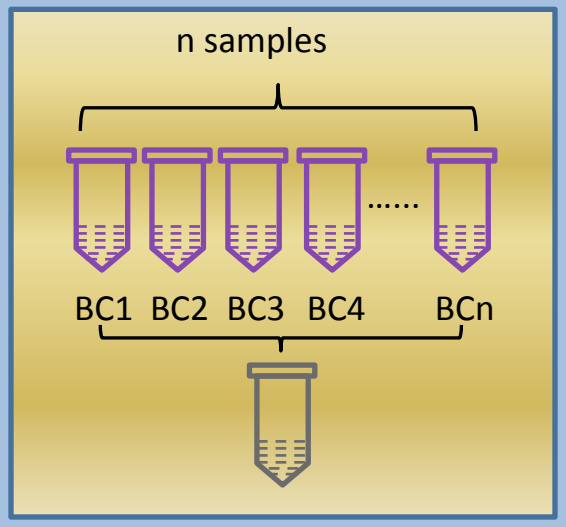
Pass Filter [Mbp]	1,500
Aligned Q17 [Mbp]	1,247
Aligned Q20 [Mbp]	1,088

Capture Sequencing at BCM-HGSC



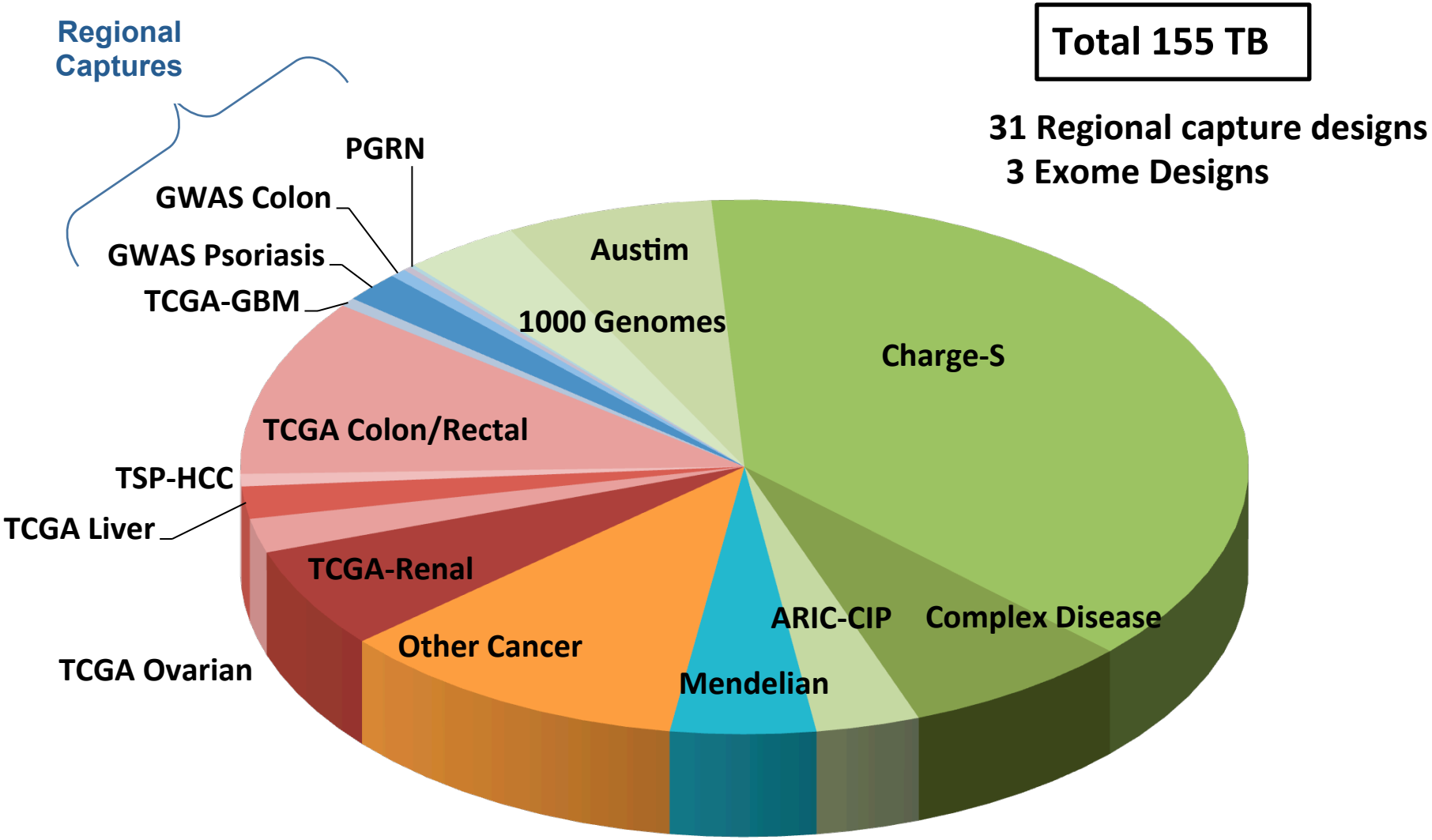
Library Automation

Decrease Reagent cost
Decrease Labor cost
Increase capture production



Multiplex Sequence Capture

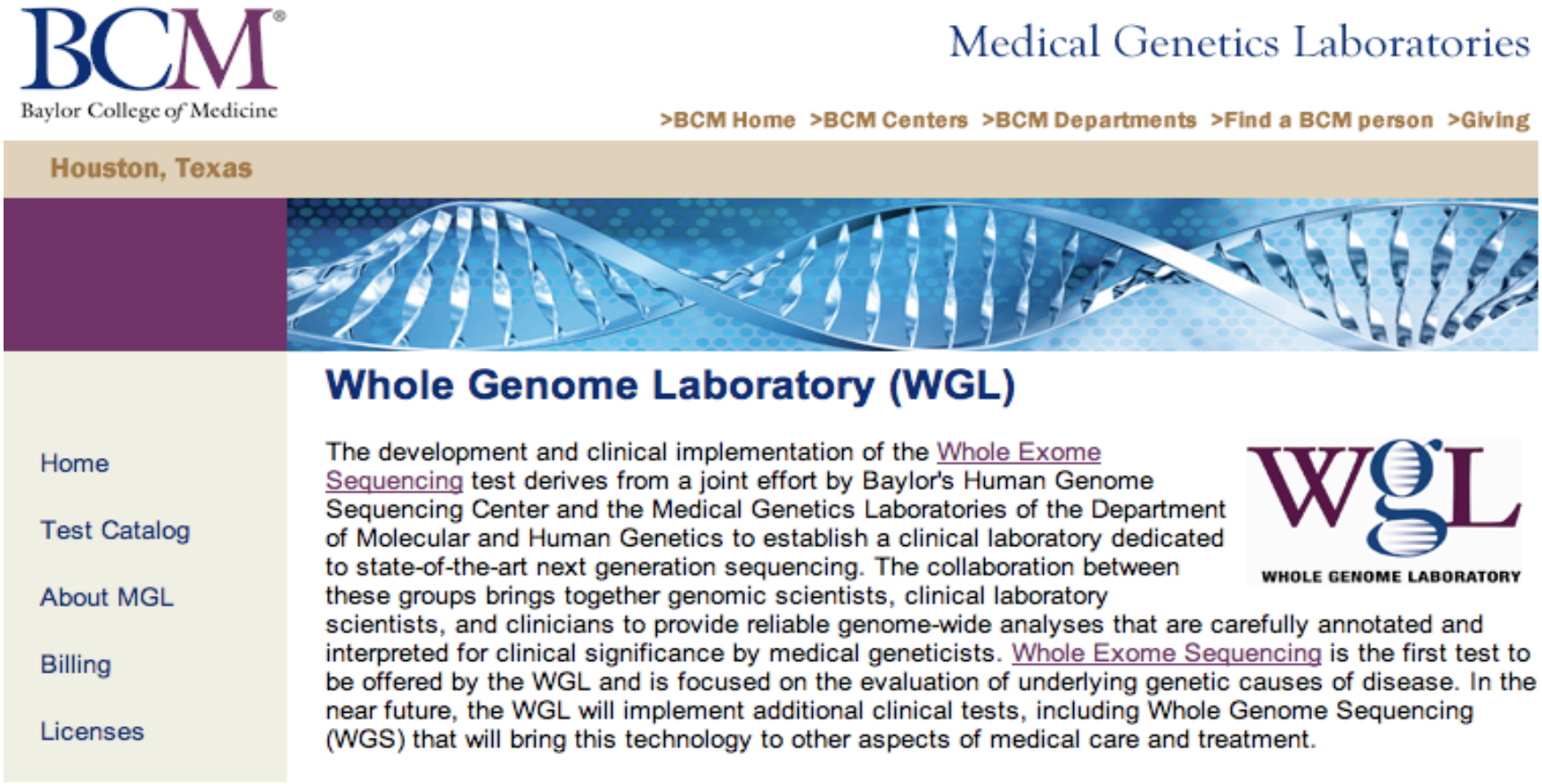
HGSC Capture Projects
May 2008 to Present



Total 155 TB

31 Regional capture designs
3 Exome Designs

Diagnostic Exome Sequencing



Joint effort between Baylor's Human Genome Sequencing Center (HGSC) and Medical Genetics Laboratories (MGL) to provide exome sequencing with clinical interpretation

Ion Torrent: Capture Applications

- Leveraged HGSC experience of Capture with SOLID and Illumina pipelines to rapidly develop the Ion Torrent capture applications.
- Library/Capture optimizations
  - Library DNA input amounts decreased to 1ug
  - Hybridization to sub-microgram amounts
  - Hybridization blocking oligo design and testing
  - Multiplexing
  - Library Automation
- Development of SNP calling tool (VarIONt) that uses Pileup data to make variant calls.

Ion Torrent: Capture Applications

Capture designs evaluated in the PGM pipeline

- All NimbleGen Liquid or Solid Probe Reagents
- Progression in capture design size and utility for diagnostic applications

Design	Genes	Targets	Genomic Region	Development
NimbleGen- chrm 8	Regional	450	116K	Library/capture dev
CHARGE-S	Regional	1.8K	2.2Mb	General control/long reads
Cancer Validation	Validation	7K	1.4Mb	TMAP/BWA comparison
Retinal panel	167	4K	1Mb	Diagnostic
Thrombosis panel	200	3K	0.5Mb	Diagnostic/Research
X Chromosome	700	7K	3Mb	Diagnostic
Cancer Exome Gapfiller	10,200	37.2K	7.4Mb	Diagnostic/Research
Exome (VCRome 2.1)	30,000	197K	43Mb	Diagnostic/Research

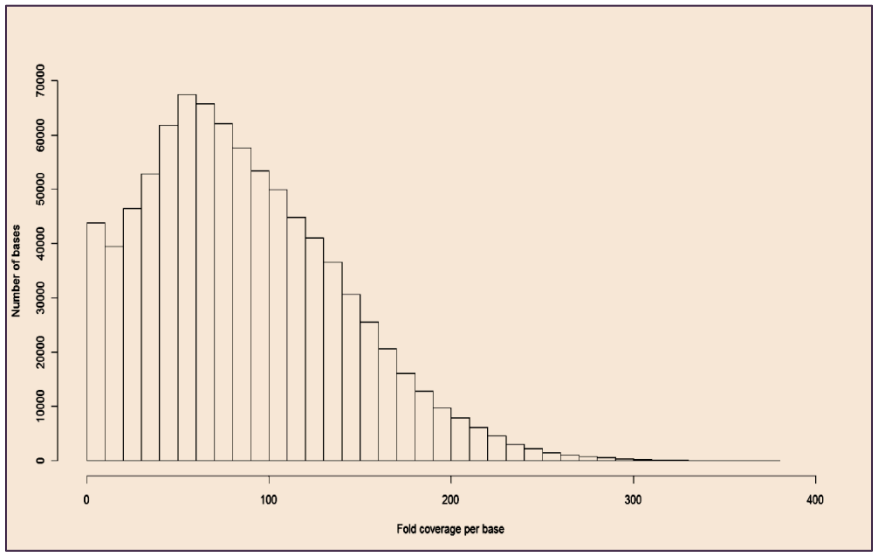
Identification of mutations in patients by PGM sequencing

Retinal Disease Penal: 167 Genes **Target design region:** 0.98Mb
Patients: Two Retinis Pigmentosa (RP) patients from two families
Sequencing: One 316 Ion Torrent Chip for each sample
Mapping: Ion Torrent TMAP

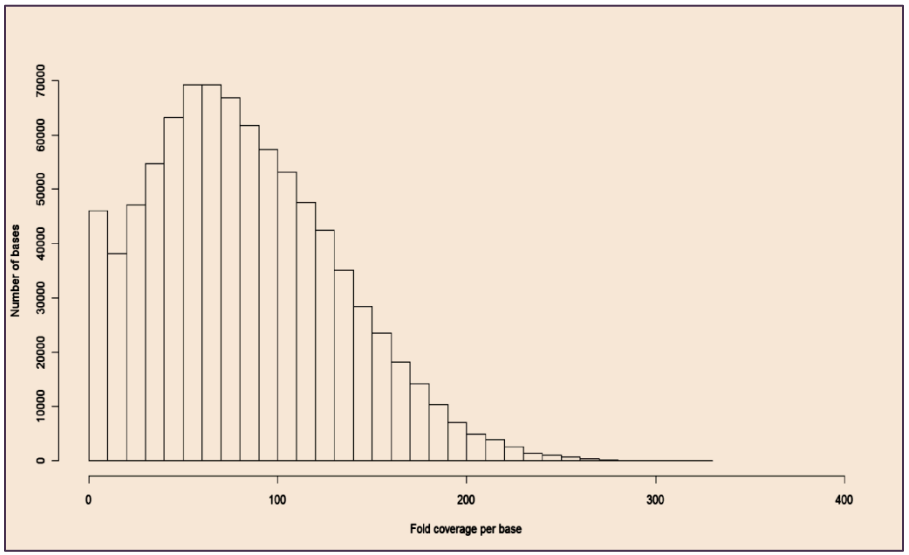
Capture Sequencing Performance of the PGM platform

Sample	Total reads	Total bases (Mb)	Mapped reads	Map ratio	In Target Ratio	Mean Coverage	% Targets Hit	% Bases 10+ coverage	% Bases 20+ coverage
Hapmap	3,487,079	391	3,211,844	92.11%	39.05%	106	99.9%	94%	90.70%
RP43	2,410,689	270	2,279,657	94.56%	46.32%	87	99.9%	94%	88.70%
RP510	2,369,797	265	2,221,340	93.74%	45.32%	84	99.9%	94%	88.70%

RP43 coverage distribution



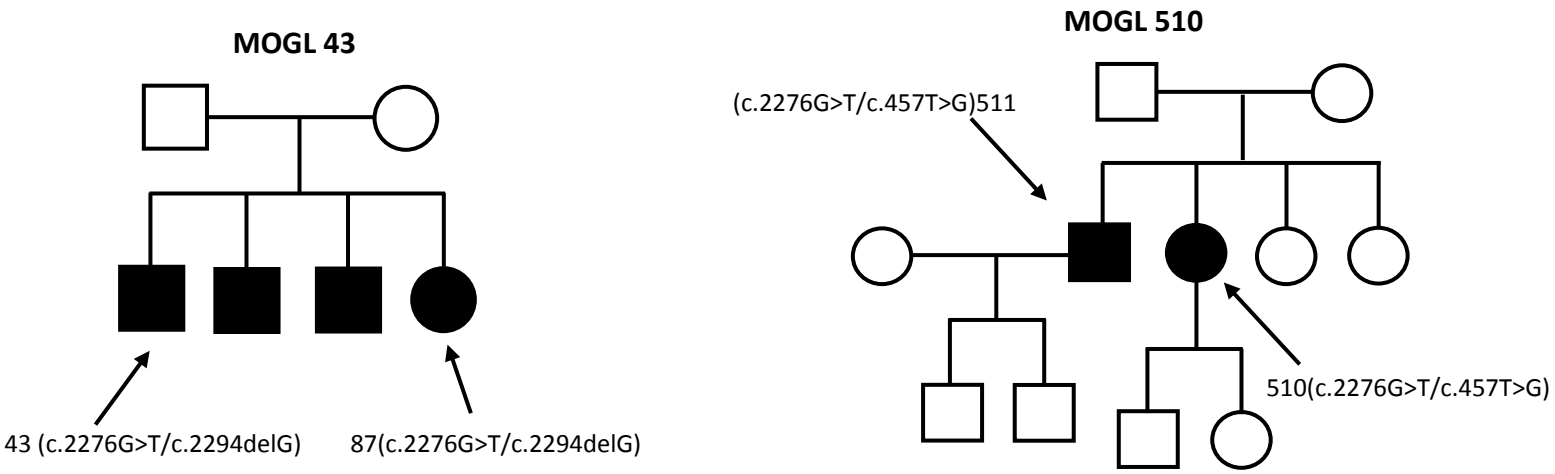
RP510 coverage distribution

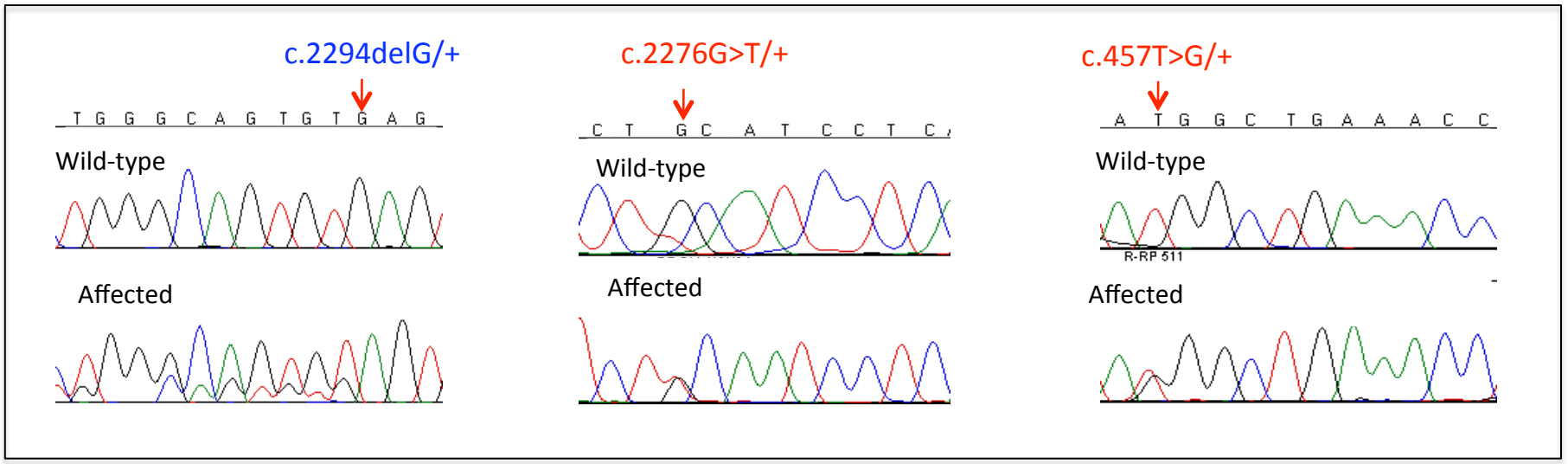


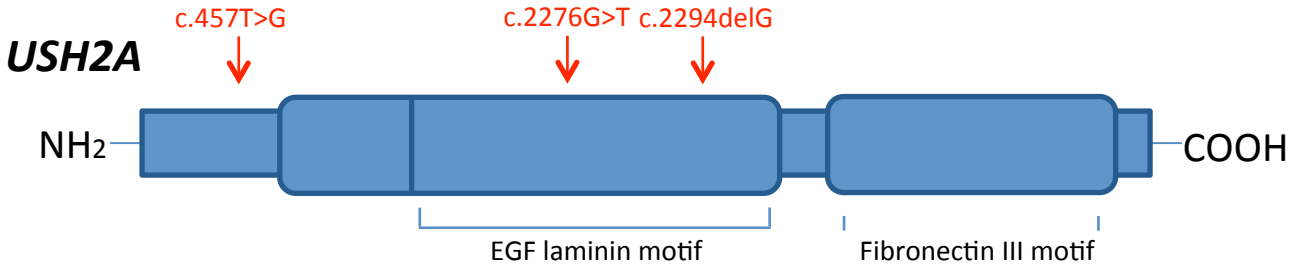
Analysis Methods

- Candidate variants, including single nucleotide variants (SNVs), insertions, and deletions (in/dels) are identified using Atlas-SNP2 and Atlas-indel
- A total of 339 and 310 SNVs are identified in coding regions for RP 43 and RP 510 respectively
- Candidate SNVs are filtered for frequently detected SNPs using dbSNP, 1000 genome database, and the HGSC internal database.
- Based on the autosomal recessive inheritance model, genes that carry at least two copies of rare coding SNVs or in/dels are identified.

- **Both patients found to carry compound heterozygous mutations in gene *USH2A***.

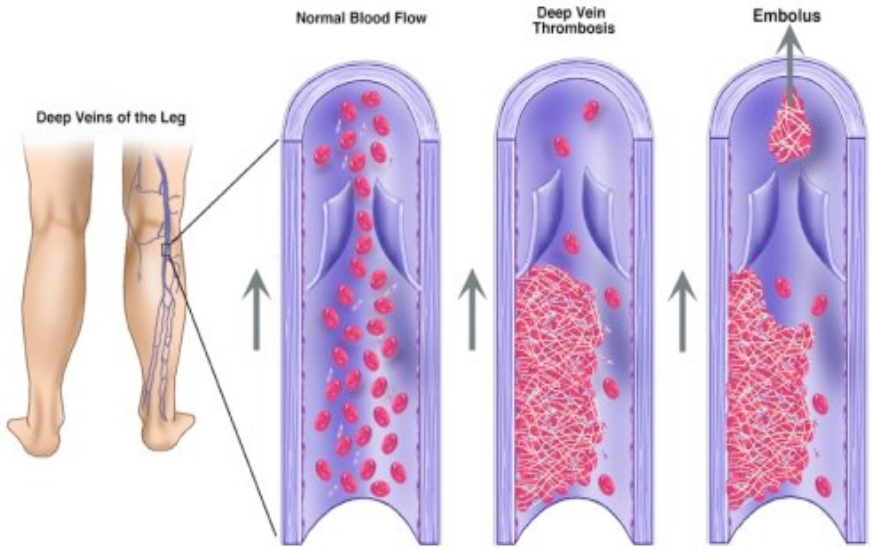

MOGL 43
43 (c.2276G>T/c.2294delG)
87(c.2276G>T/c.2294delG)
MOGL 510
(c.2276G>T/c.457T>G)511
510(c.2276G>T/c.457T>G)


c.2294delG/+
T G G G C A G T G T G A G
Wild-type
Affected
c.2276G>T/+
C T G C A T C C T C
Wild-type
Affected
c.457T>G/+
A T G G C T G A A A C C
Wild-type
R-RP 511
Affected


USH2A
c.457T>G
c.2276G>T c.2294delG
NH2
COOH
EGF laminin motif
Fibronectin III motif

Definition of the phenotype: deep vein thrombosis (DVT)

- Definition: Pathological formation of a blood clot within the lumen of a vein in deep circulation (usually affects lower limbs)
- May be idiopathic (no precipitating cause) or secondary (trauma, surgery, cancer, etc.)



Deep Vein Thrombosis (DVT)

Established DVT genetic risk variants:

- **Natural anticoagulant deficiencies:** Include rare loss of functions mutations in **SERPINC1, PROC and PROS1** leading to deficiencies of natural anticoaglant protiens (antithrombin, protein C, protein S).

  Deficiencies are detected biochemically, by measuring the functional activity of these proteins in plasma

- **Factor V Leiden and Pro-thrombin G20210A (FII)**
  Mechanism of association: hypercoagulability
  (MAF = ~4% in general population).

- **Other common variants known to be associated with DVT**:
  - chrX:138460946; DVT-risk allele A
  - chr4:155727040; DVT-risk allele C
  - chr4:187357205; DVT-risk allele C

- Established variants only explain a fraction of disease heritability

Ion Torrent Regional Capture: Deep Vein Thrombosis (DVT)

Thrombosis Design: ~200 Genes; **Target design region:** 0.50Mb.
Patients: 20 Patients diagnosed with DVT
Capture/Sequencing: Samples were barcoded and multiplexed capture.
Sequenced using 318 chips.

Mapping: Ion Torrent TMAP

Capture Sequencing Performance of the PGM platform

Sample	Aligned (Mb)	% Duplicate Reads	% Total Reads Aligned	% Aligned Reads On-Target	Average Coverage	% Reads on target or buffer	% Targets Hit	% Bases with 10+ coverage	% Bases with 20+ coverage
5108-0	408	28%	100%	15%	99	16%	95%	85%	81%
5769-0	424	26%	100%	15%	106	16%	94%	81%	77%
7781-0	315	29%	99%	13%	67	14%	95%	82%	77%
486-7	286	31%	100%	18%	85	18%	94%	77%	73%
Ion Torrent Full Length Hyb Blockers (8 plex capture)									
16 Samples	187	26%	98%	42%	141	45%	98%	96%	95%

Analysis Methods

- Reads mapped using Ion Torrent TMAP
- The resulting BAM file was subjected to Samtools Pileup and HGSC VarIONt caller for SNV discovery specifically designed to work with Ion Torrent data (*See Christian Buhay, Talk and Poster*).
- Candidate SNVs are filtered for frequently detected SNPs using dbSNP, 1000 genome database the HGSC internal database and HGMD as part of the Mercury annotation pipeline. *(See David Sexton, Mercury Analysis and Annotation Pipeline)*
- On average, we discovered ~320 SNVs per patient in exons across the Thrombosis chip. The fraction of non-synonymous variants is roughly 40% of all variants found (or 130 per patient).
- Examined Known Thrombosis (DVT) Genes and Risk Factors.

Thrombosis: Analysis Results

Thrombosis Variant Overview				
DVT Risk Alleles and Gene Targets	Patient Samples			
	5108-0	5769-0	7781-0	486-7
SERPINC1				X
Factor V Leiden	X	X	X	
chrX DVT risk allele	X	X		
chr4 DVT risk allele		X		X
chr4 DVT risk allele	X	X	X	X

Samples and variants sequenced by Ion Torrent matched expected risk profiles.
- Patient **486-7** confirmed anti-thrombin deficiency by functional tests -- Regional capture and analysis found 2 novel, rare, non-syn variants found in **SERPINC1**

- Confirmed Factor V Leiden allele in Patients **5108-0, 5769-0 and 7781-0**. Originally determined by Sanger sequencing

- Identified at least one known common risk SNP for DVT in each of the patients in addition to the other already known risk factors

Exome Capture Applications

Whole exome data using Ion Torrent's PGM. Two exomes were sequenced: one hapmap sample (NA12763), and a patient with Charcot-Marie-Tooth Neuropathy (HS1011). Both samples have been sequenced previously with Illumina and have SNP-array data.

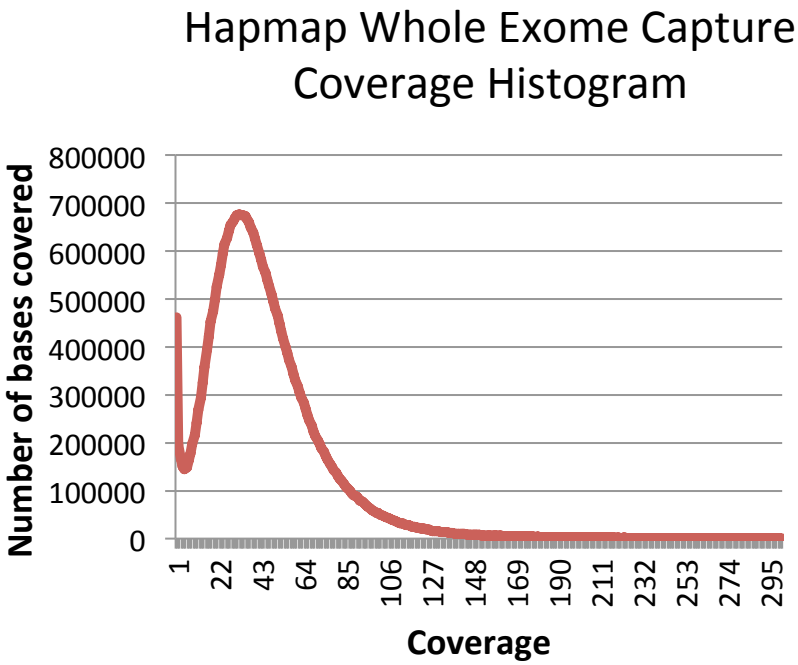
Sequenced: Multiple runs using Ion Torrent's 318 chip were merged to achieve (~7Gb).

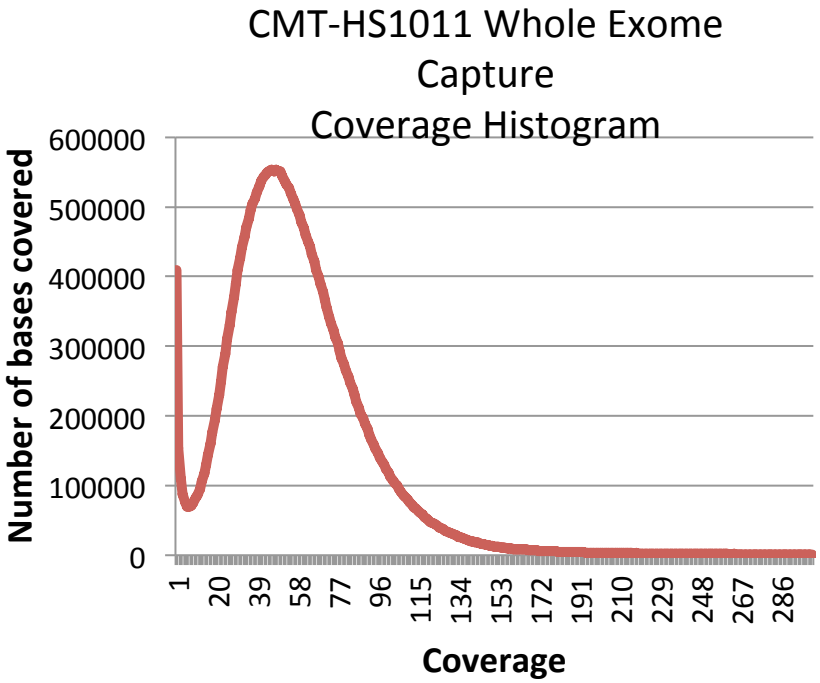
Exome Capture: HGSC VCRome 2.1 design; 30K genes; 43 Mb target regions
Mapping: Ion Torrent TMAP

Exome Capture Results: ≥84% of target bases were covered at 20X or better.
Full Length Hybridization Blocking Oligos improves "On Target Performance"

Sample	Chip Type	Runs Generated	Aligned (Gb)	% Duplicate Reads	% Total Reads Aligned	Average Coverage	% Reads on target or buffer	% Targets Hit	% Bases10+ coverage	% Bases 20+ coverage
NA12763	318	10	6.8	10%	98%	43	42%	99.30%	94.50%	84.00%
HS1011	318	10	7.4	12%	99%	57	40%	99.50%	96.60%	92.60%
Exome Capture with Full Length Hyb Blocking Oligos										
HS1011	318	10	6.3	27%	92%	74	59%	99.25%	96.23%	94%

Exome Analysis





Variant Analysis utilized Pileup and the HGSC VarIONt caller. Variant calls from were compared to the SNP array data in the case of the HapMap NA12763 sample resulting in 99.1% concordance. Variants for the CMT Sample were compared to the initial Illumina exome data resulting in 92% SNP concordance and the **2 causative mutations identified in *SH3TC2* as previously reported in NEJM.**

Summary/Conclusions

- Successfully implemented the Ion Torrent PGM for production applications.
- Demonstrated platform utility for Capture applications in discovery and validation (Retinal Disease, Thrombosis and Exome Sequencing).
- Successfully demonstrated the performance of the platform to detect pathogenic mutations.
- Continue to harden the Ion Torrent capture and analysis pipelines for application in a clinical environment.

Acknowledgements

- Richard Gibbs
- Eric Boerwinkle
- Jeff Reid
- David Wheeler

Ion Torrent Sequencing Group
  - Christian Buhay**
  - Mike Holder
  - Huyen Dinh
  - Christie Kovar
  - Imad Khalil
  - Lora Lewis
  - Bob Ruth

Retinal Disease Gene Project
  - Rui Chen
  - Xia Wang**

Thrombosis Disease Gene Project
  - Luca Lotta (Milan, Italy)

Library Capture Group
  - Mark Wang**
  - Chunmei Qu

NimbleGen
  - Tom Albert
  - Jeffrey Jeddeloh
  - Daniel Burgess

Acknowledgements

Ion Torrent/Life Technologies

- Kelly Hoon
- Tim Harkins
- Marcella Putman
- Jason Myers
- Ilya Zlatkovsky
- Jessica Reed
- Anup Parikh

Ion Torrent Regional Capture: Retinal Disease Genes

(Project performed in Collaboration with Xia Wang and Rui Chen)

Retinal Gene Design: 167 Genes

Target design region: 0.98Mb

Initial Test Sequencing: HapMap Sample NA11831 from CEPH panel

Sequencing: Total of 19 Ion Torrent runs: 17 run of 314 Chips and two runs on 316 chips

Read Mapping: Ion Torrent-TMAP

Sample	Total reads	Total bases (Mb)	Mapped reads	Map ratio	In Target Ratio	Mean Coverage	% Targets Hit	% Bases 10+ coverage	% Bases 20+ coverage
Hapmap	3,487,079	391	3,211,844	92.11%	39.05%	106	99.9%	94%	90.70%

Affy SNP 6.0 array concordance

1185 genotyped bases overlap

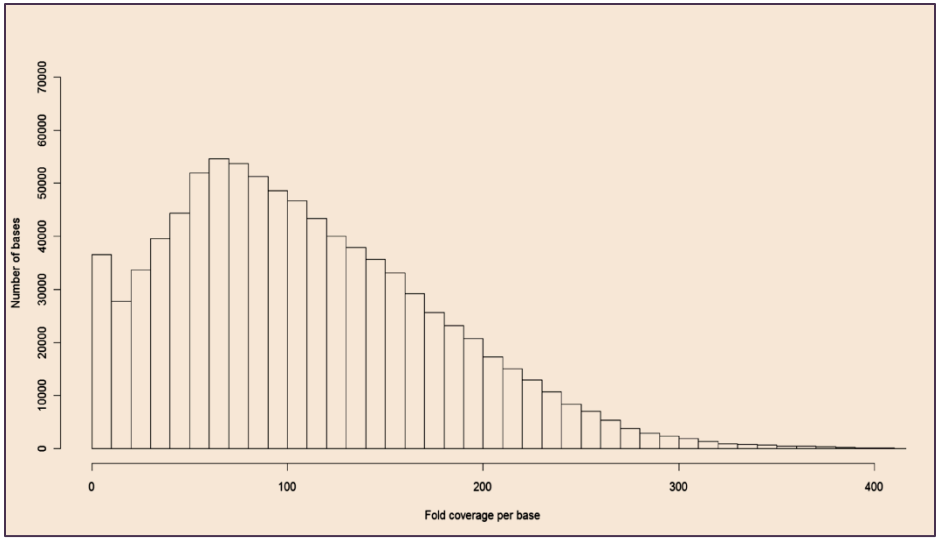
Accuracy: 99.59%

Sensitivity: 99.24%

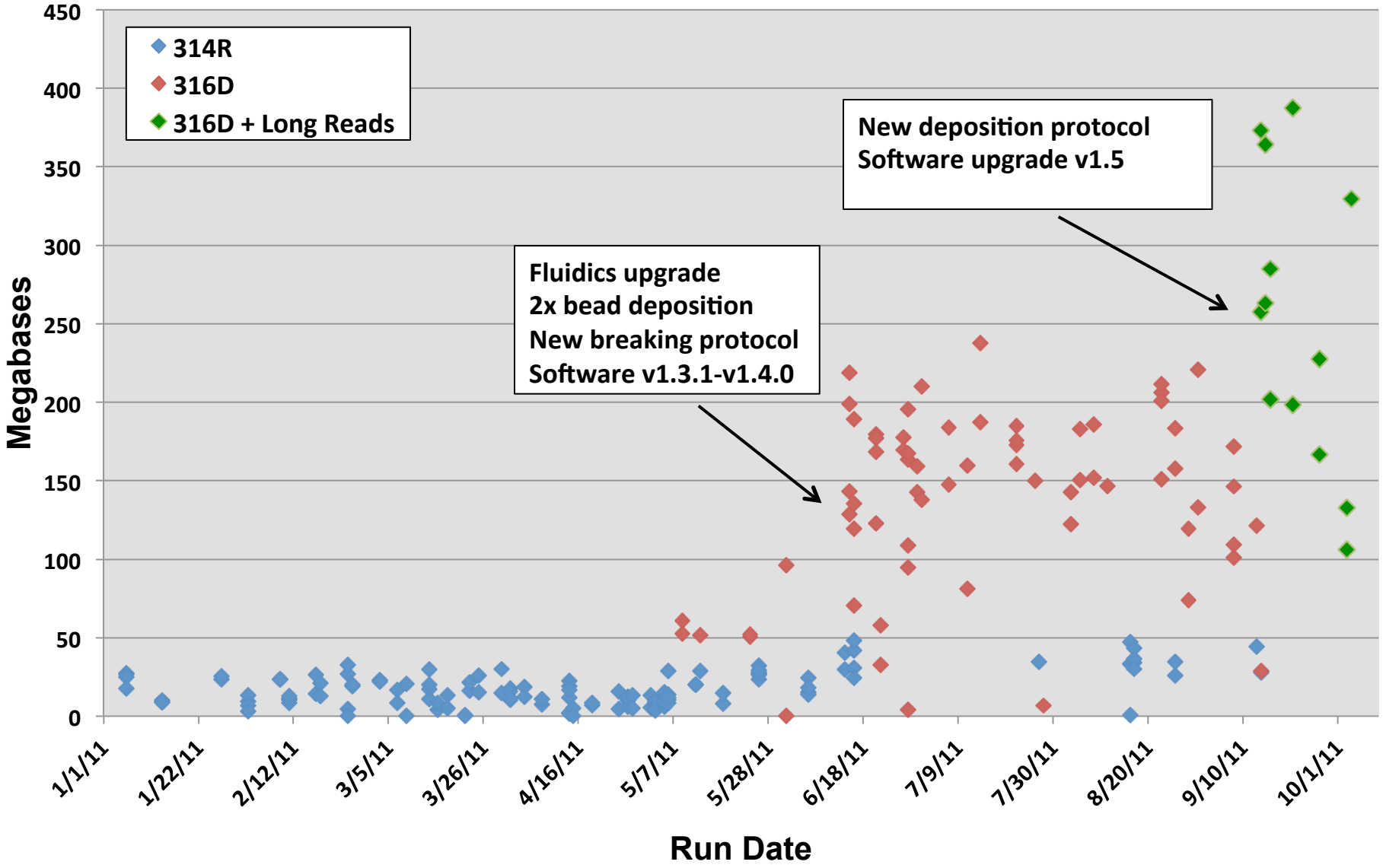
Specificity: 99.67%

Total 5 discordant SNPs

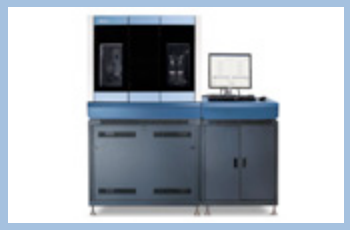
NA11831 coverage distribution



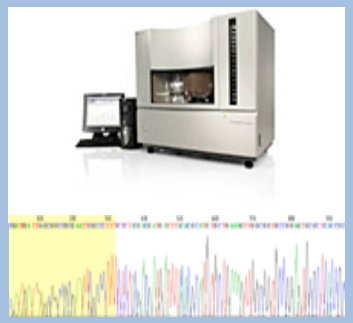
Ion Torrent Production
Aligned Q17 Mb/Run



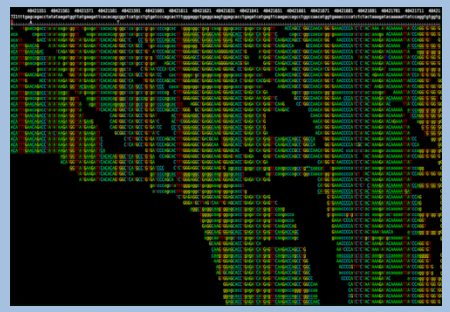
BCM-HGSC Sequencer Fleet

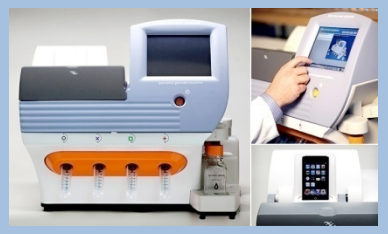


Life Tech SOLiD™ 4



Life Tech 3730 system





Life Tech Ion Torrent PGM

Roche 454 FLX system

Illumina GAIIx & Hiseq2000

Pacific Biosciences

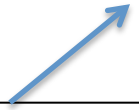
Ion Torrent Regional Capture: Deep Vein Thrombosis (DVT)

Thrombosis Design: ~200 Genes; **Target design region:** 0.50Mb.
Patients: 20 Patients diagnosed with DVT
Capture/Sequencing: Samples were barcoded and multiplexed capture. Sequenced using 318 chips.

Mapping: Ion Torrent TMAP

Capture Sequencing Performance of the PGM platform

Sample	Aligned (Mb)	% Duplicate Reads	% Total Reads Aligned	% Aligned Reads On-Target	Average Coverage	% Reads on target or buffer	% Targets Hit	% Bases with 10+ coverage	% Bases with 20+ coverage
5108-0	408	28%	100%	15%	99	16%	95%	85%	81%
5769-0	424	26%	100%	15%	106	16%	94%	81%	77%
7781-0	315	29%	99%	13%	67	14%	95%	82%	77%
486-7	286	31%	100%	18%	85	18%	94%	77%	73%
Ion Torrent Full Length Hyb Blockers (8 plex capture)									
16 Samples	187	26%	98%	42%	141	45%	98%	96%	95%



Full Length Hybridization Blocking Oligos improves "On Target Performance"

Identification of mutations in patients by PGM sequencing

Retinal disease Panel on one member from two independent families diagnosed with retinitis pigmentosa (RP)

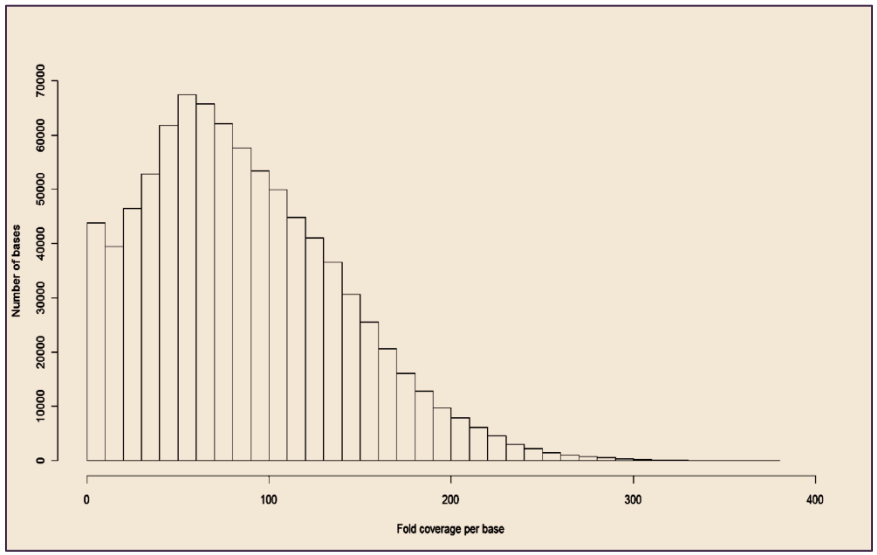
Sequencing: One 316 Ion Torrent Chip for each sample

Mapping: Ion Torrent TMAP

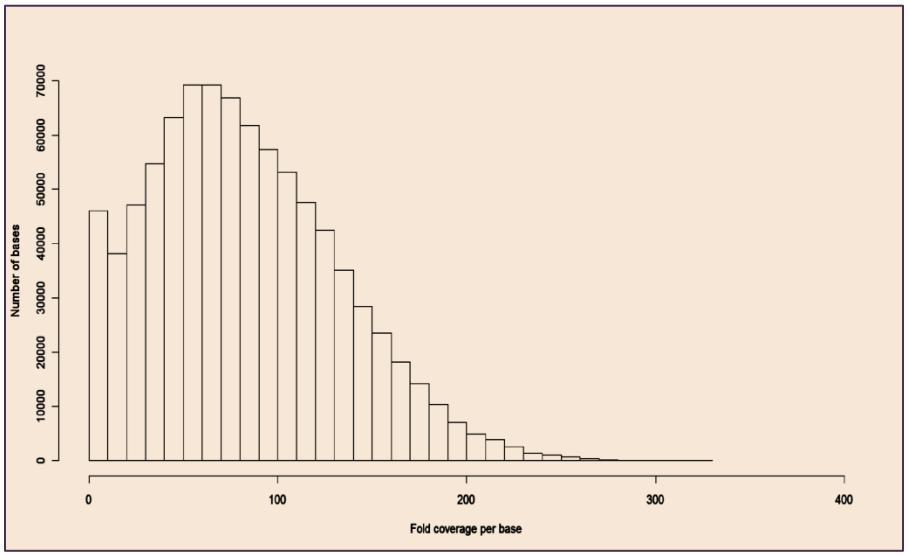
Sequencing performance of hapmap and patients samples in PGM platform

Sample	Total reads	Total bases (Mb)	Mapped reads	Map ratio	In Target Ratio	Mean Coverage	% Targets Hit	% Bases 10+ coverage	% Bases 20+ coverage
Hapmap	3,487,079	391	3,211,844	92.11%	39.05%	106	99.9%	94%	90.70%
RP43	2,410,689	270	2,279,657	94.56%	46.32%	87	99.9%	94%	88.70%
RP510	2,369,797	265	2,221,340	93.74%	45.32%	84	99.9%	94%	88.70%

RP43 coverage distribution



RP510 coverage distribution



Thrombosis: Analysis Results

Thrombosis Variant Overview				
DVT Risk Alleles and Gene Targets	Patient Samples			
	5108-0	5769-0	7781-0	486-7
Factor V Leiden	X	X	X	
chrX DVT risk allele	X	X		
chr4 DVT risk allele		X		X
chr4 DVT risk allele	X	X	X	X
SERPINC1				X

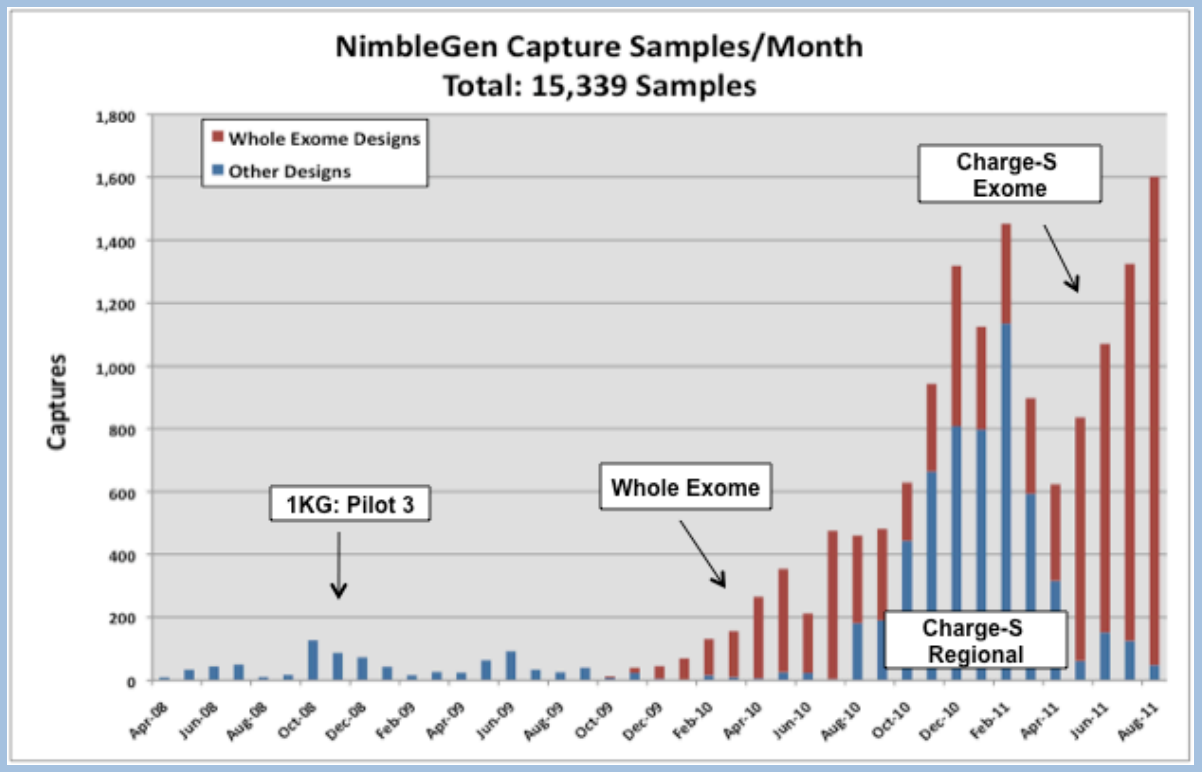
Samples and variants sequenced by Ion Torrent matched expected risk profiles.

-Patient **486-7** was identified to have anti-thrombin deficiency by functional tests --> **SERPINC1** gene. Regional capture and analysis found 2 novel, rare, non-syn variants found in SERPINC1

-Confirmed Factor V Leiden allele in Patients **5108-0, 5769-0 and 7781-0**. Originally determined by Sanger sequencing

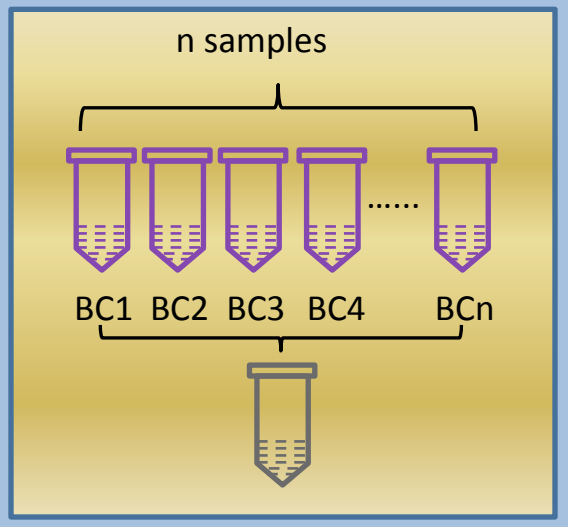
-Identified at least one known common risk SNP for DVT in each of the patients in addition to the other already known risk factors

Capture Sequencing at BCM-HGSC



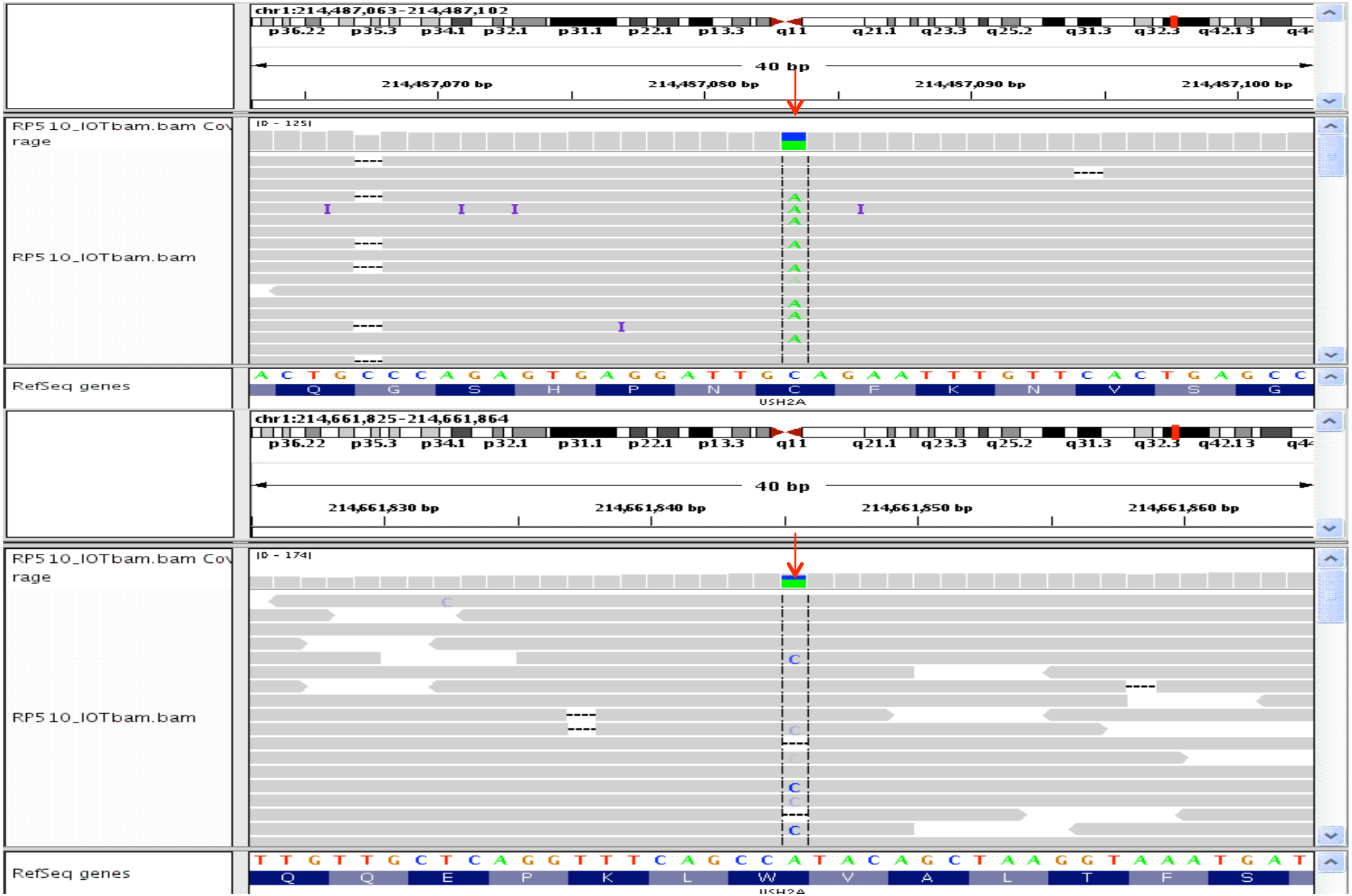
Library Automation

Decrease Reagent cost
Decrease Labor cost
Increase capture production

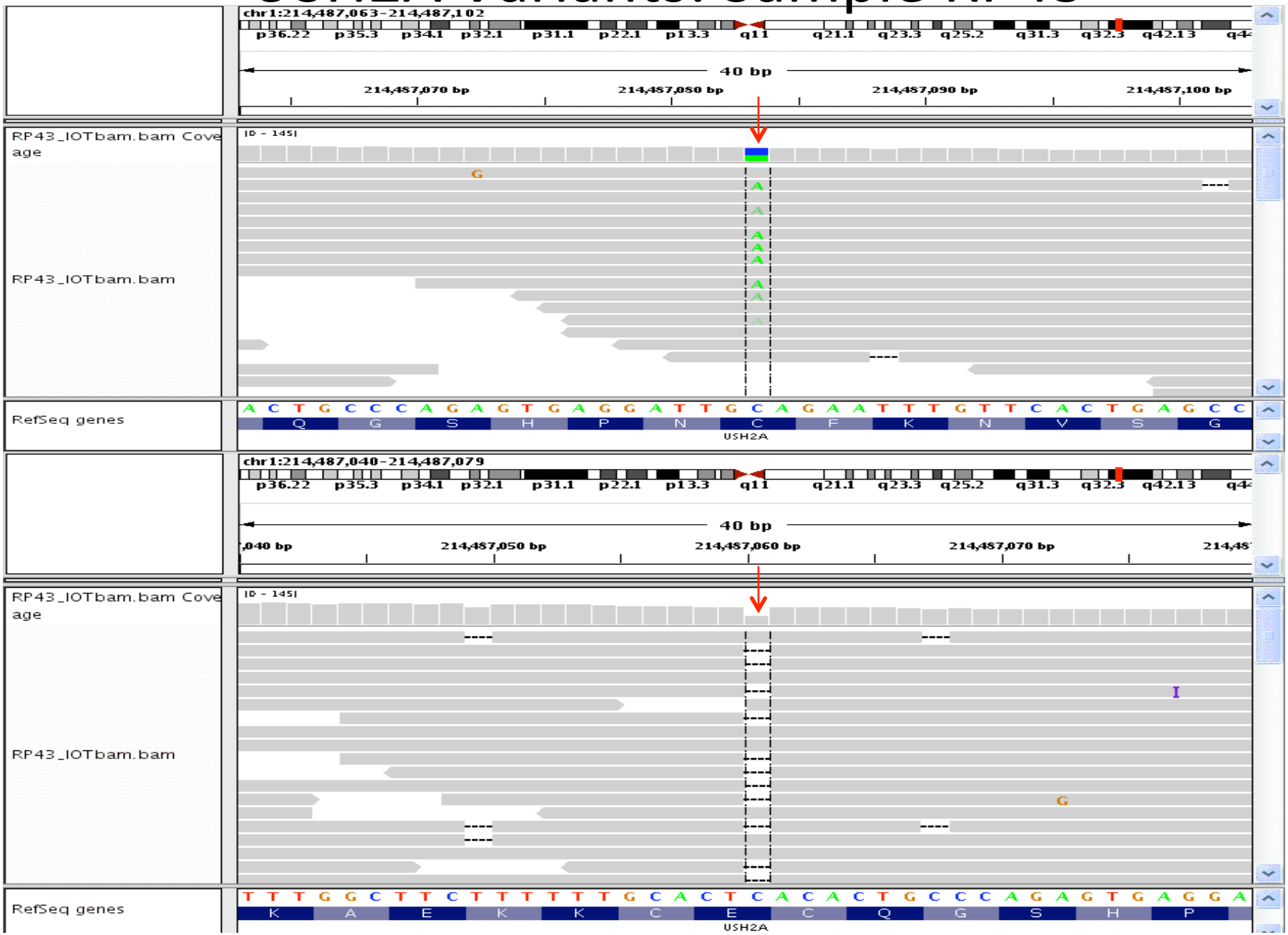


Multiplex Sequence Capture

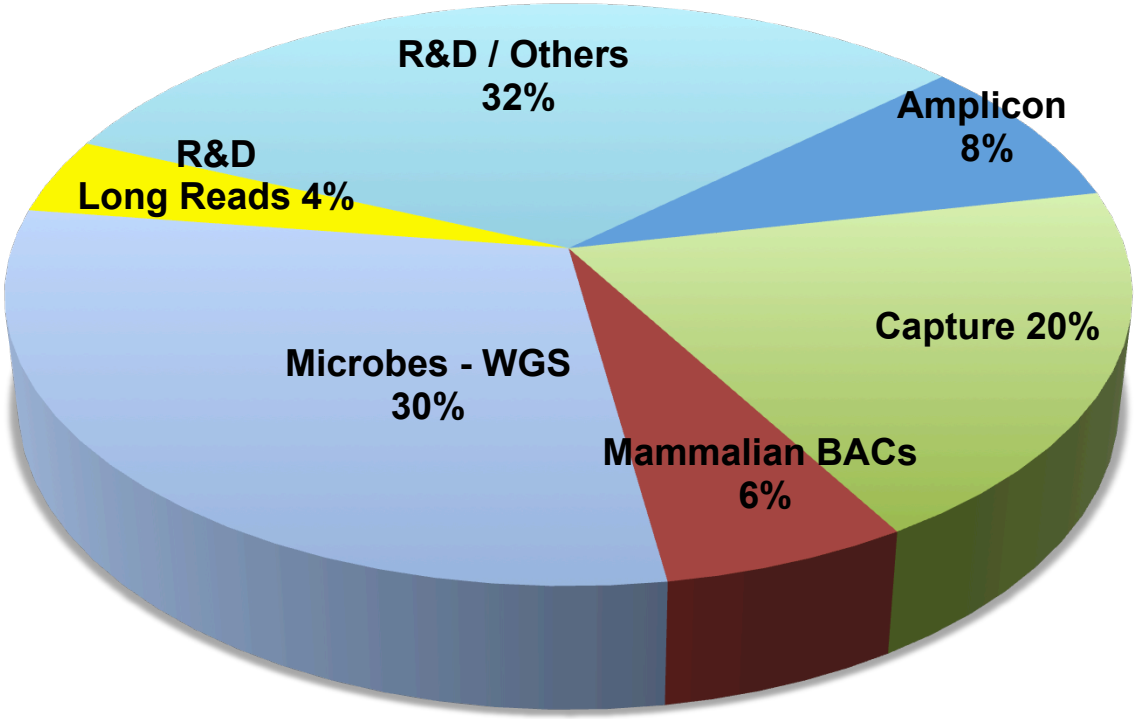
USH2A Variants: Sample RP510



USH2A Variants: Sample RP43



Distribution of Ion Torrent Runs



Total Runs	**213**
Total Gb	**20.6**

Current Metrics - 20 runs

Average 316 Run + Long Reads

Pass Filter Mb/Run	373
Q17 Aligned Mb/Run	253
Q20 Aligned Mb/Run	204
Pass Filter RL	205
Q17 Aligned RL	158
Q20 Aligned RL	136

Best In House 316 + Long Reads

Pass Filter Mb/Run	519
Q17 Aligned Mb/Run	387
Q20 Aligned Mb/Run	316
Pass Filter RL	224
Q17 Aligned RL	176
Q20 Aligned RL	156

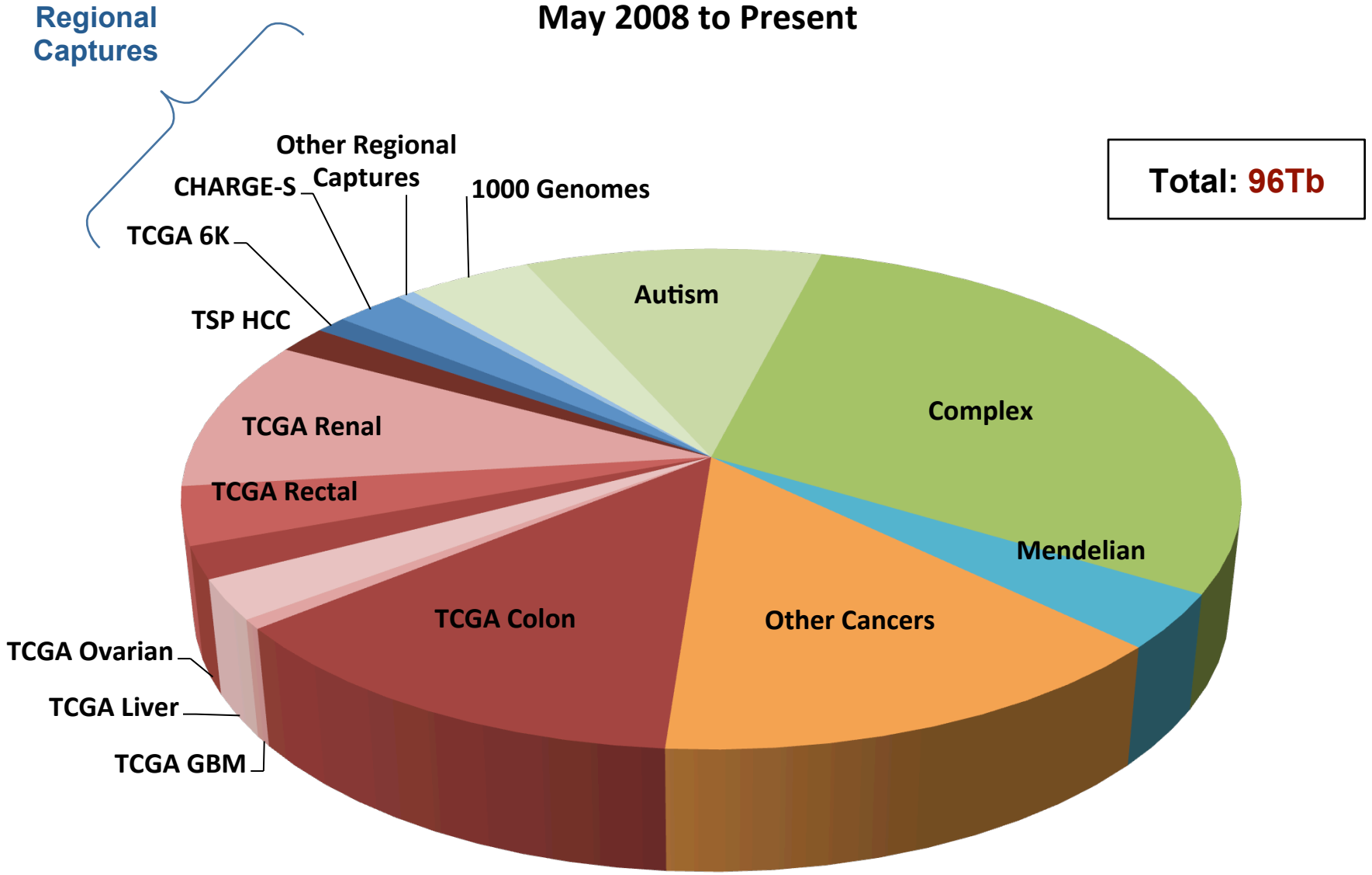
Analysis Methods

- Retinis Pigmentosa (RP) follows an autosomal recessive inheritance pattern as both male and female are affected and neither parents of the affected members show RP phenotype.
- Candidate variants, including single nucleotide variants (SNVs), insertions, and deletions (in/dels) are identified using Atlas-SNP2 and Atlas-indel
- A total of 339 and 310 SNVs are identified in coding regions for RP 43 and RP 510 respectively
- Candidate SNVs are filtered for frequently detected SNPs using dbSNP, 1000 genome database, and the HGSC internal database.
- Based on the autosomal recessive inheritance model, genes that carry at least two copies of rare coding SNVs or in/dels are identified.

- **Both patients found to carry compound heterozygous mutations in gene *Ush2A***.

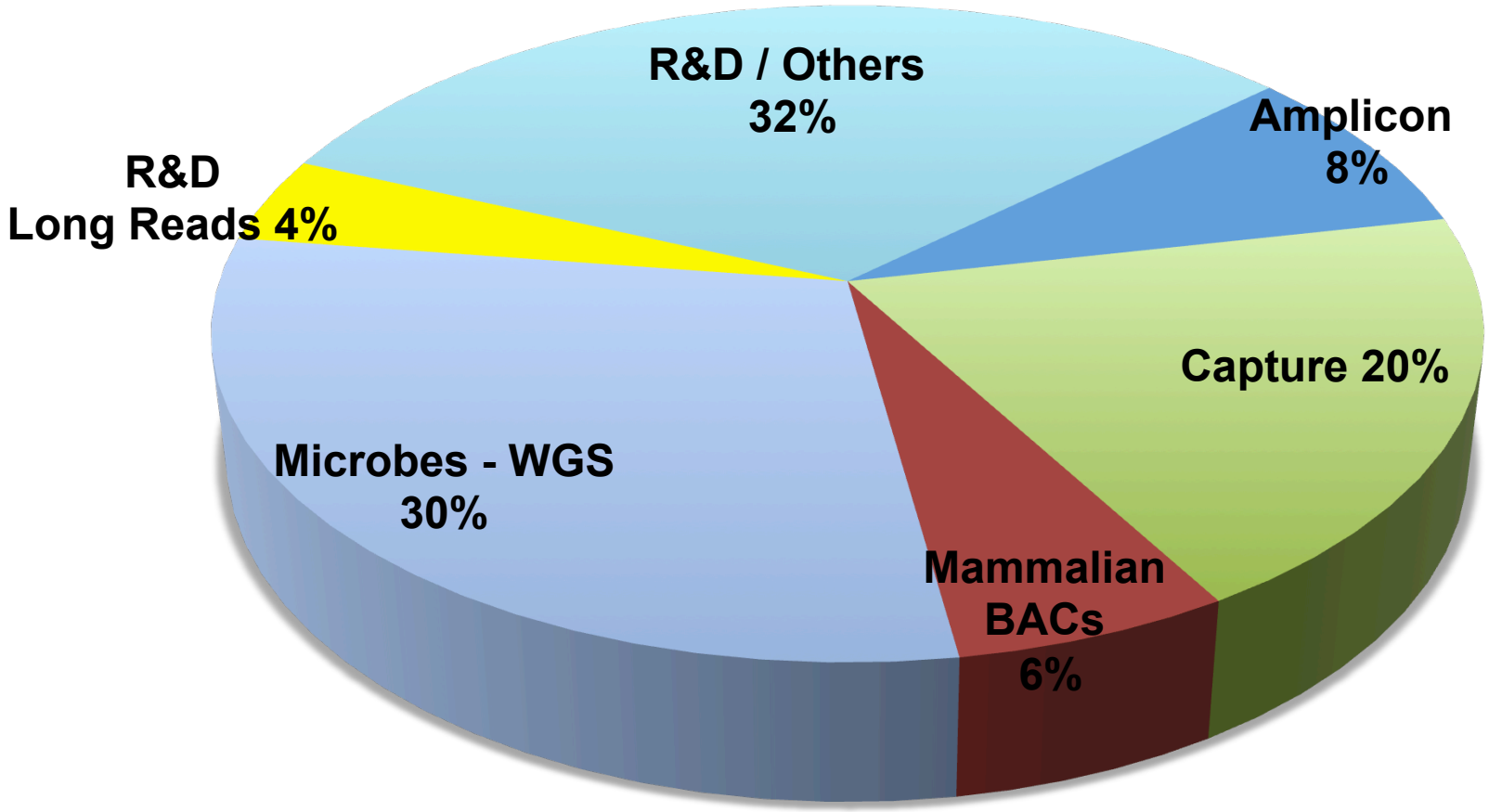
HGSC Capture Projects

May 2008 to Present



Distribution of Ion Torrent Runs

Total Runs 213
Total Gb 20.6



Ion Torrent: Capture Applications

- Five regional designs evaluated in the PGM pipeline
- All NimbleGen liquid Probe Reagents

Design	Genes	Targets	Genomic Region	Development
CHARGE-S	Regional	1.8K	2.2Mb	General control/long reads
Cancer Validation	Validation sites	7K	1.4Mb	TMAP/BWA comparison
Retinal panel	167	4K	1Mb	Diagnostic
NimbleGen- chrm 8	Regional	450	116K	Library/capture dev
NimbleGen- chrm 7	Regional	440	112K	Library/capture dev
TCGA exome Gapfiller			7Mb	

- Library/Capture optimizations
  - Library DNA input amounts decreased to 1ug
  - Hybridization to sub-microgram amounts
  - Hybridization blocking oligo design and testing
- Evaluation of Ion Torrent TMAP and BWA mapping tools

Amplicon Sequencing Methods and Applications

Two methods

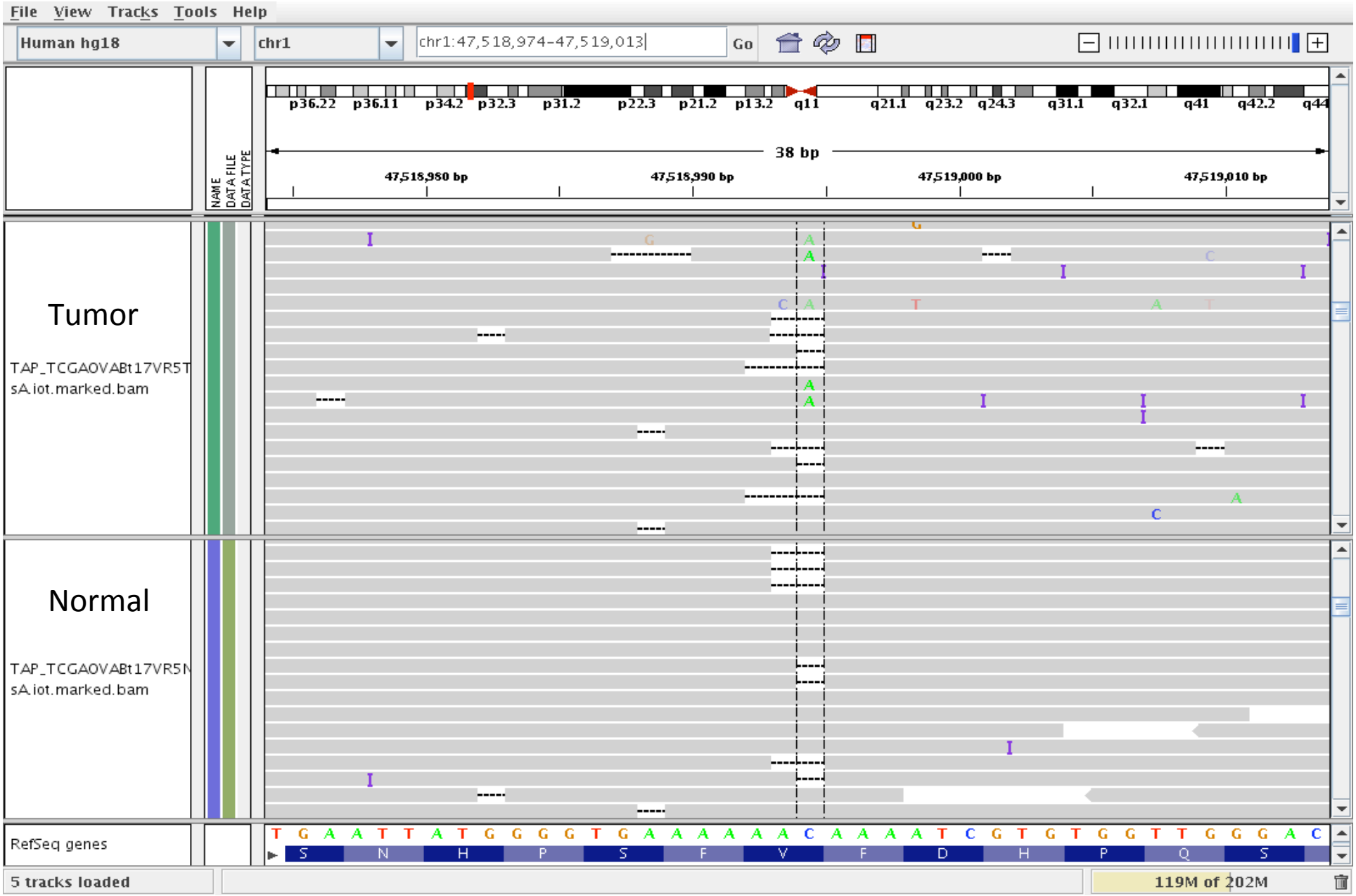
- Modification of amplicon design pipeline to target 100-150bp amplicons

- iShear protocol for larger amplicons >500bp – enzymatic shear
  - Allows for use of existing amplicon sets and comparison

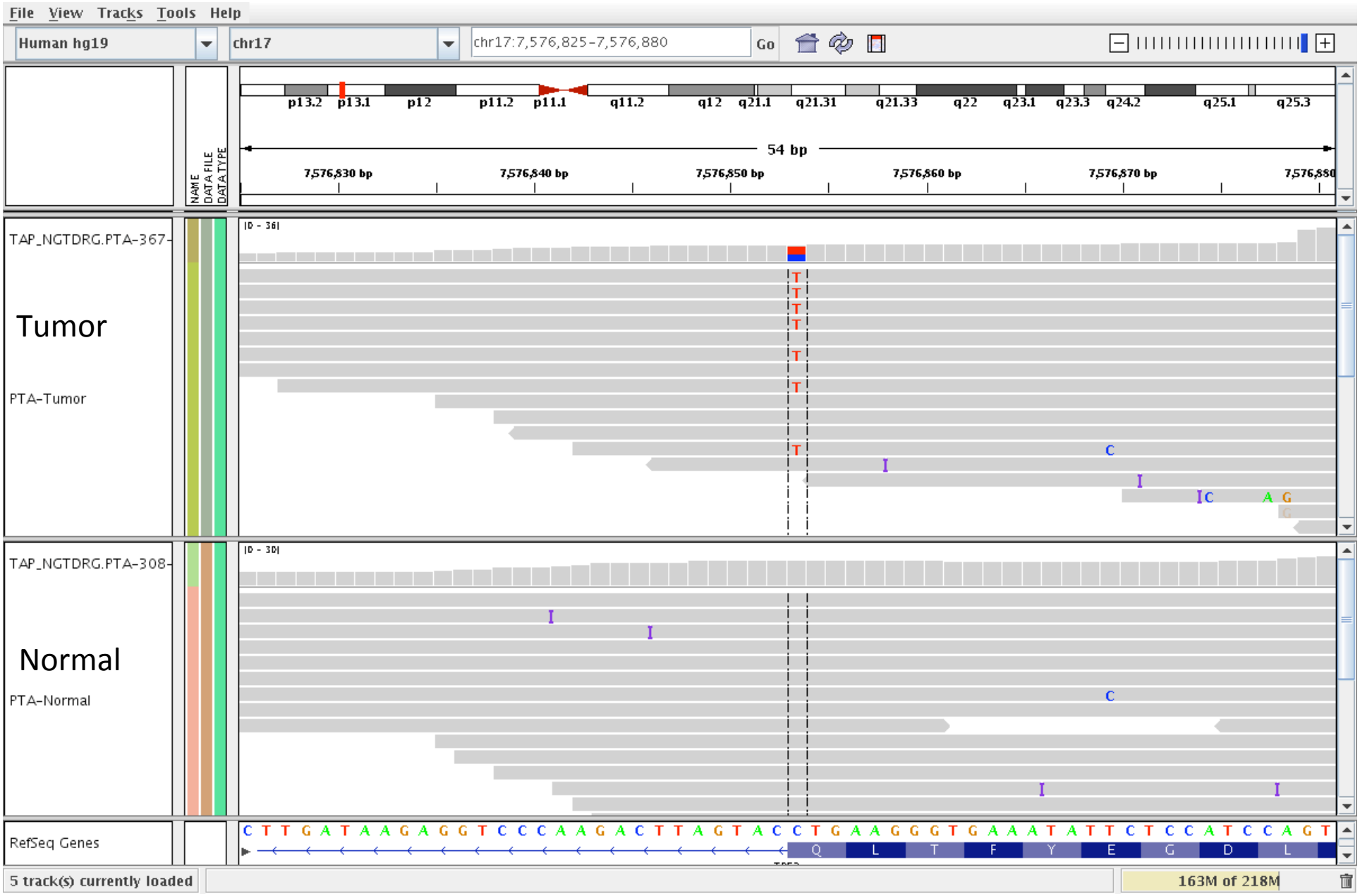
Summary of Amplicon Sequencing Activities

Design	Application	Amplicons	Total size	Development
Cancer 4 Gene Set	Discovery	166	8.8 Kb	Initial amplicon testing
1000 Genomes	Validation sites	289	33 Kb	316 Chip evaluation/iShear
TCGA Ovarian	Validation sites	931	154 Kb	200bp reads/iShear
TCGA Colon/Rectal Set 1	Validation sites	1256	250 Kb	200bp reads
TCGA Colon/Rectal Set 2	Validation sites	1400	250 Kb	200bp reads

Confirmed validation site – Visual Validation

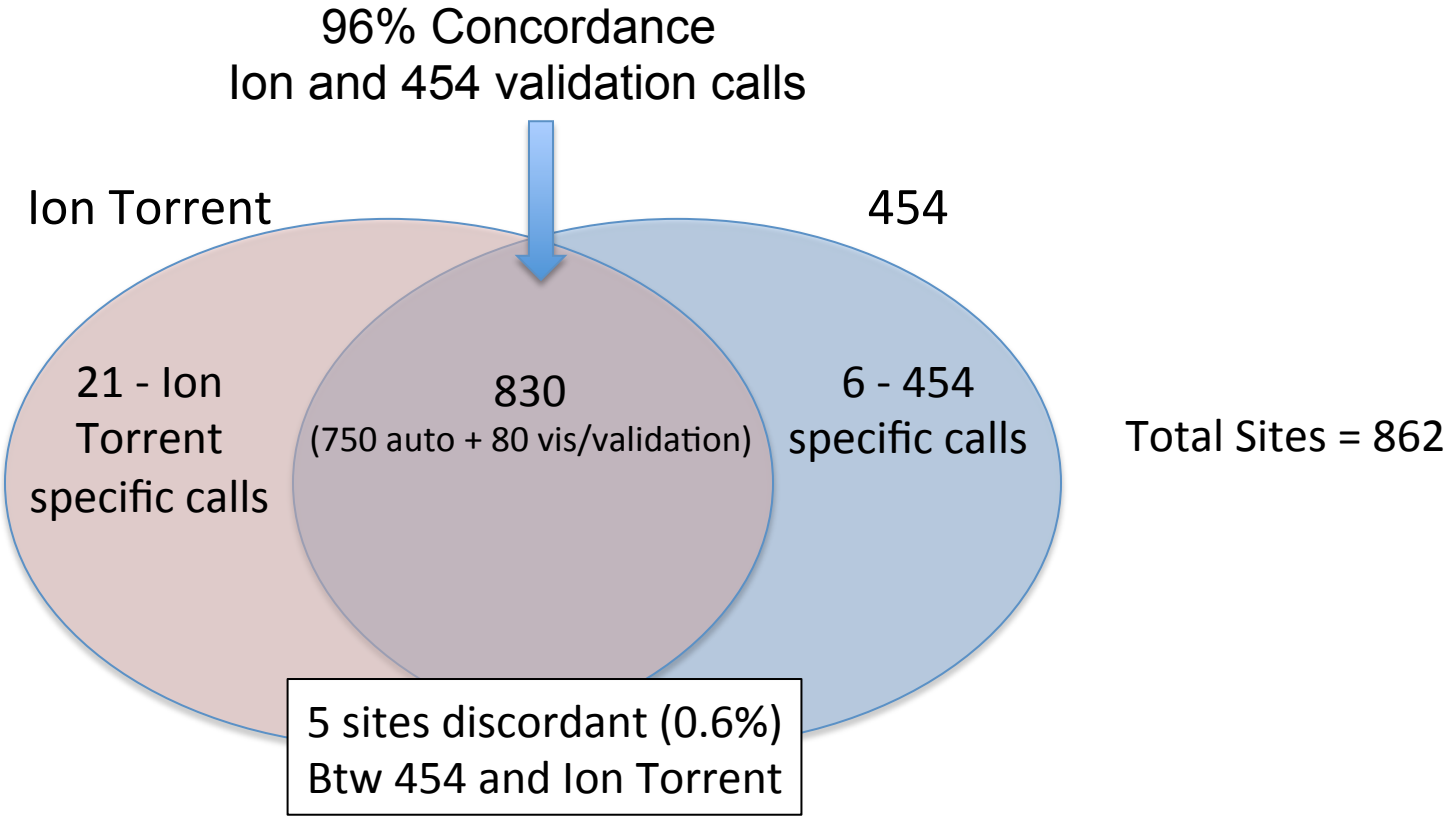


Example of Confirmed Validation Site

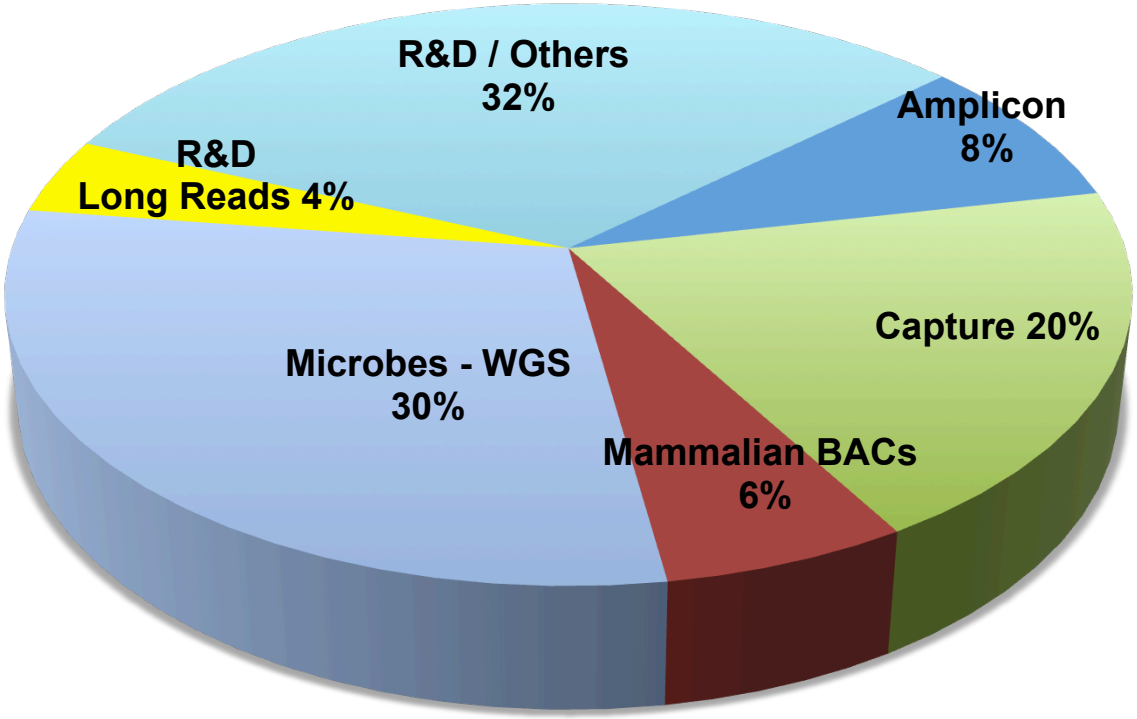


Comparison of Validation Results

- **Mapping:** Ion Torrent TMAP
- **Variant calling pipeline:** Samtools Pileup.
  - Require ≥ 5% variant allele frequency for positive variant detection (consistent with 454 requirements).
- **Methods:** include automated methods followed by visual validation for any sites discordant between 454 and Ion Torrent.



Distribution of Ion Torrent Runs



Total Runs	**213**
Total Gb	**20.6**

Current Metrics - 20 runs

Average 316 Run + Long Reads

Pass Filter Mb/Run	373
Q17 Aligned Mb/Run	253
Q20 Aligned Mb/Run	204
Pass Filter RL	205
Q17 Aligned RL	158
Q20 Aligned RL	136

Best In House 316 + Long Reads

Pass Filter Mb/Run	519
Q17 Aligned Mb/Run	387
Q20 Aligned Mb/Run	316
Pass Filter RL	224
Q17 Aligned RL	176
Q20 Aligned RL	156

Amplicon Sequencing Methods and Applications

Two methods

- Modification of amplicon design pipeline to target 100-150bp amplicons

- iShear protocol for larger amplicons >500bp – enzymatic shear
  - Allows for use of existing amplicon sets and comparison

Summary of Amplicon Sequencing Activities

Design	Application	Amplicons	Total size	Development
Cancer 4 Gene Set	Discovery	166	8.8 Kb	Initial amplicon testing
1000 Genomes	Validation sites	289	33 Kb	316 Chip evaluation/iShear
Ovarian	Validation sites	931	154 Kb	200bp reads/iShear
Colon/Rectal Set 1	Validation sites	1400	250 Kb	200bp reads: in Progress
Colon/Rectal Set 2	Validation sites	1400	250 Kb	200bp reads: in Progress

Variant Validation: Amplicon Sequencing

- **TCGA Ovarian Validation Set:**
  - SOLiD exome sequencing for discovery and validated on 454
- **Ion Torrent Amplicon pool:**
  - 931validation sites; 862 SNV's and 69 indel sites
- **Total Design size:** 154Kb
- **Sequencing:** iShear library protocol; Ion 316 chip + 200bp reads
- **Avg coverage:** Tumor/Normal at least 850X with 98.5% of bases covered ≥20X

Sample	PF Run Yield (Mb)	Est. Aligned Yield (Mb)	% Total Reads Aligned	Average Coverage	% Reads that hit target or buffer	% Targets Hit	% Bases with 1+ coverage	% Bases with 10+ coverage	% Bases with 20+ coverage
Ova-Tumor	259	203	99.14%	909	94%	99%	99%	98.61%	98.50%
Ova-Norm	251	204	97.70%	867	94%	99%	99%	98.54%	98.53%

Ion Torrent: Capture Applications

- Five regional designs evaluated in the PGM pipeline
- All NimbleGen liquid Probe Reagents

Design	Genes	Targets	Genomic Region	Development
CHARGE-S	Regional	1.8K	2.2Mb	General control/long reads
Cancer Validation	Validation sites	7K	1.4Mb	TMAP/BWA comparison
Retinal panel	167	4K	1Mb	Diagnostic
NimbleGen- chrm 8	Regional	450	116K	Library/capture dev
NimbleGen- chrm 7	Regional	440	112K	Library/capture dev

- Leveraged HGSC experience of Capture with SOLID and Illumina pipelines to rapidly develop the Ion Torrent capture applications.
- Library/Capture optimizations
  - Library DNA input amounts decreased to 1ug
  - Hybridization to sub-microgram amounts
  - Hybridization blocking oligo design and testing
- Evaluation of Ion Torrent TMAP and BWA mapping tools